Math 10B with Professor Stankova
Worksheet, Discussion #29; Thursday, 5/9/2019
GSI name: Roy Zhao

1 Continuous Random Variables

1. True False If the mean doesn't exist, then the standard deviation doesn't exist.
2. True False If the mean exists, then the standard deviation exists.
3. True False CDFs are always continuous but PDFs don't have to be.
4. True False If we know the CDF F, we find the probability $P(2 \leq X \leq 5)$ without calculating the PDF f.
5. True False The variance of a symmetric random variable centered at 0 is $\int_{-\infty}^{\infty} x^2 f(x)dx$.
6. True False If a is the median of a continuous random variable X with PDF f, then $\int_a^{\infty} f(x)dx = \frac{1}{2}$.
7. Let X be a random variable on a probability space Ω with a probability function P and let f be the PDF for X, F the CDF. Draw a picture of how all these variables interact and explain any special arrows that you have in your diagram.
8. Let $g(x) = \begin{cases} x^2 & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$. Find c such that $f(x) = cg(x)$ is a PDF. Graph f and the CDF F. Find the mean and median of $f(x)$.
9. Let $F(x) = \frac{x-1}{x+1}$ for $x \geq 1$ and 0 for $x \leq 1$. Show that F is a CDF. Find the PDF associated with it and the probability that we choose a number between 1 and 2.
10. Let $f(x)$ be $-2x$ from $-1 \leq x \leq 0$ and 0 everywhere else. Find the standard deviation of this distribution.
11. Let $f(x) = e \cdot e^x$ for $x \leq -1$ and 0 otherwise. Find the standard deviation of this distribution.
12. Prove all of the formulas for mean and variance for each of the random variables below:

Distribution	PDF	$E(X)$	Variance
Uniform	$f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
Exponential	$f(x) = ce^{-cx}$	$\frac{1}{c}$	$\frac{1}{c^2}$
Laplacian	$f(x) = \frac{1}{2}e^{-\|x\|}$	0	2

2 CLT

13. True False A smaller 95% confidence interval means that we are less sure about what the mean μ could be.

14. True False The normalized random variable $Z = \frac{\bar{X}-\bar{\mu}}{\sigma/\sqrt{n}}$ is standard normally distributed.

15. Suppose that the height of women is distributed with an average height of 63 inches and a standard deviation of 10 inches. Taking a sample of 100 women, what is the probability that the average of the heights of these 100 women is between 62 and 64 inches?

16. Suppose the weight of newborns is distributed with an average weight of 8 ounces and a standard deviation of 1 ounce. Today, there were 25 babies born at the Berkeley hospital. What is the probability that the average weight of these newborns is less than 7.5 ounces?

17. I have a (possibly biased) coin and flip it 100 times and get heads 90 times. What is the 95% confidence interval for p, the probability of flipping a heads?

18. Assume the standard deviation of student heights is 5 inches. How large of a sample do you need to be 95% confident that the sample mean is within 1 inch of the population mean?

19. In a class of 25 students, the time that students spent on the midterm was 40 minutes with a standard deviation of 5 minutes. What is the 95% confidence interval for the average time taken on the midterm?

20. I have a loaded die and I think that it is more likely to be a 1 than normal. Suppose I roll it 100 times and get 1 25 times. What is the 95% confidence interval for p, the probability of getting a 1?

21. Every day, the number of people who are born is Poisson distributed with an average of 4900 people per day. We count how many people are born in a span of 100 days and let $\bar{X}$ denote the average number of people born per day. What is the probability $P(\bar{X} \leq 4895)$?

3 Hypothesis Testing

22. True False We never accept the null hypothesis, we just fail to reject it.

23. True False The Z test only works for random variable $X_1, \ldots, X_n$ normally distributed.

24. True False The T test only works for random variable $X_1, \ldots, X_n$ normally distributed.

25. True False The T test only works for $n < 30$.

26. Which type of hypothesis test should you use for the following situations: (PMF Hypothesis Test/Z-Test/T-Test/χ^2 Goodness-of-fit/χ^2 Independence)

    - You want to know if a die is fair so you roll it 100 times and count the number of $1, 2, \ldots, 6$s
    - You want to know if a die is biased towards 5 so you roll it 100 times
    - You want to know if a student's letter grade is related to whether they go to office hours
    - You want to know if a coin is biased towards heads and flip it 10 times
    - You want to know if flipping coins give you a binomial distribution so you ask 500 friends to flip a coin 5 times and count the number of friends who flipped $0, 1, \ldots, 5$ heads
    - You want to know if a coin is biased towards heads and flip it 100 times
    - You want to know if course evals are independent of the section so you count the number of $1, 2, \ldots, 7$s per section
    - You know heights are normally distributed and want to know if Berkeley students are taller than normal so you take 10 student heights

27. (True story) A woman claims that she can smell when someone has Parkinson's disease. She is given 10 people's shirts and correctly said whether the person had the disease in 9 of the 10 cases. Does she have this ability with $\alpha = 0.05$? (The 10th person who she said had Parkinson's actually developed it months later so she was really 10 for 10).

28. When counting families with 2 children, I find that 83 of them have two girls, 102 of them have two boys, and 215 of them have one boy and one girl. Suppose that my null hypothesis is having a boy or girl is that I expect a $1 : 1 : 2$ ratio. Can we reject the null hypothesis with $\alpha = 0.05$?

29. The height of 4 NBA players is 75 inches with a sample standard deviation of $s = 6$. Can we reject the null hypothesis that NBA players are the same height as the average person's height (which is 66 inches) with a two sided alternative hypothesis and $\alpha = 0.05$?

30. Write the PDF for the t distribution when $\nu = 4$.

31. You think that students who go to office hours will do better than those that don't. Suppose that exam scores are normally distributed with $\mu = 70$. You ask 9 students who go to office hours what their exam scores were and get a sample mean of $\bar{x} = 74$ and a sample standard deviation of $s = 6$. Can you say that students that go to office hours do better than other students with $\alpha = 0.05$?

32. Find $\chi^2(x)$ for $k = 1, 2, 4$.

33. Every year 25% of people contract the flu. This year, the NIH comes out with a vaccine and out of 1600 people, there are only 350 people who contract the disease. Was the vaccine successful with $\alpha = 1\%$?

34. You are wondering whether performing well in this course and gender are related and you get the following table. Are they related?

	Male	Female
Pass	315	485
Fail	85	115

35. In a skittle bag, you get 14 red skittles, 12 blue, 1 green, 10 yellow, and 13 orange skittles. Is it possible that the colors are evenly distributed with a significance level of $\alpha = 0.05$?

4 MLE

36. True False For MLE, setting $\frac{df}{dx} = 0$ will always give a minimum or maximum of f.

37. True False The estimator you get from maximum likelihood estimation will be unbiased.

38. There is a bag with 10 balls colored red and blue. You pull out two balls (with replacement) and get BR. What is the maximum likelihood for the number of blue balls in bag?

39. You have a coin that you think is biased. you flip it 3 times and get the sequence HTH. What is the maximum likelihood estimate for the probability of getting heads?

40. You assume that the lifespan of lightbulbs are exponentially distributed (PDF is $\lambda e^{-\lambda t}$ for $t \geq 0$) and notice that your three light bulbs go out in 1, 2, and 3 years. What is the maximum likelihood estimator for λ?

41. I go to Kip's and want to figure out the total number of students n there. By looking, I see that I've taught 2 of the students there. I pick a student at random and it turns out to be one of the students I've taught. Then 8 more students come in and of those I have taught 7 of them. Now I again pick a random student and I haven't taught this second student picked. What is the most likely number n of total students at Kip's originally?

5 Miscellaneous

42. True False The line of best fit is the line that minimizes the least square error.

43. Suppose that three people randomly pick a hat. What is the expected value of the number of people who choose their hat? (with proof). What is the variance? Now do the same with n people.

44. Stirling's approximation tells us that $n! \approx \frac{n^n\sqrt{2n\pi}}{e^n}$. Use this to show that $\binom{2n}{n} \approx \frac{2^{2n}}{\sqrt{n\pi}}$.

45. Prove that $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$.

46. Use induction to prove that $\Gamma(n) = (n-1)!$ for all $n \geq 1$.

47. Use induction to prove that $E[\chi^2_{k=2n}(x)] = 2n$ for all $n \geq 1$.

48. What is the definition of variance? Prove that $Var(X) = E[X^2] - E[X]^2$.

49. What is the definition of covariance? Prove that $Cov(X,Y) = E[XY] - E[X]E[Y]$.

50. The formulas for the slope and y intercept of the line of best fit come from MLE. Suppose that error is normally distributed. This means that if we predict $y = ax_i + b$, then the probability of actually getting y_i follows the PDF

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-(y_i-y)^2/2\sigma^2} = \frac{1}{\sigma\sqrt{2\pi}}e^{-(y_i-(ax_i+b))^2/2\sigma^2}.$$

Use MLE to show that $\hat{b} = \bar{y} - a\bar{x}$.

51. Now with $b = \bar{y} - a\bar{x}$, do MLE to show that $\hat{a} = r\frac{\sigma_y}{\sigma_x}$ the formula that we use for a.